

BEST AVAILABLE COPY

PCT/JP 2004/003838

日 本 国 特 許 庁
JAPAN PATENT OFFICE

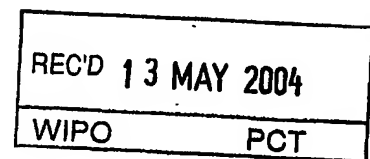
22. 3. 2004

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: 2003年 3月22日

出 願 番 号
Application Number: 特願2003-120733
[ST. 10/C]: [JP 2003-120733]



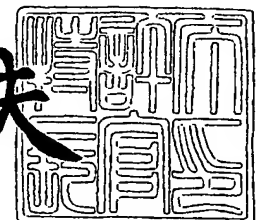
出 願 人
Applicant(s): 株式会社サン・フレア

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2004年 4月23日

特許庁長官
Commissioner,
Japan Patent Office

今 井 康 夫



【書類名】 特許願
【提出日】 平成15年 3月22日
【整理番号】 P0150322
【あて先】 特許庁長官 殿
【国際特許分類】 G09B
【発明者】

【住所又は居所】 東京都国分寺市富士本 2 丁目 2 9 番 3 号

【氏名】 徳田 尚之

【発明者】

【住所又は居所】 カナダ国、V 2 N 6 X 7 プリティッシュ コロンビア
、 プリンス ジョージ、3 9 0 1 ウィーブ ロード
、 ユニット 2 1 6

【氏名】 陳 亮

【特許出願人】

【識別番号】 596091749

【住所又は居所】 東京都新宿区四谷 4 丁目 7 番地

【氏名又は名称】 株式会社サン・フレア

【代理人】

【識別番号】 100082153

【弁理士】

【氏名又は名称】 小原 二郎

【選任した代理人】

【識別番号】 100080078

【弁理士】

【氏名又は名称】 駒津 敏洋

【手数料の表示】

【予納台帳番号】 004710

【納付金額】 21,000円

【提出物件の目録】

【物件名】 明細書 1

【物件名】 要約書 1

【物件名】 図面 1

【包括委任状番号】 9723424

【書類名】 明細書

【発明の名称】 テンプレートーテンプレート構造に基づく新しい対話システム

【特許請求の範囲】

【請求項 1】 抽出ルール、ベースのテンプレートーテンプレート構造およびバギー・ルールの拡張性を用いた対話式学習システム。

【請求項 2】 テンプレートーテンプレート構造内のノードのいくつかは、多くのテンプレートまたはいわゆる大きいテンプレートに抽出することができる抽出ルールに関連付けられたシンボルでマーク付けされるテンプレートーテンプレート構造。

【請求項 3】 展開テンプレートのノードのいくつかは抽出ルールに関連付けられた特別なシンボルが割り当てられており、これによってテンプレートの 1 つまたは複数を抽出することができるようになされたテンプレートーテンプレート・スキーム。

【請求項 4】 単一テンプレートーテンプレートから複数のテンプレートを抽出するためにテンプレートーテンプレート内のノードをマーク付けするために使用される抽出ルール。

【請求項 5】 スリムなテンプレートーテンプレートをより大きなテンプレートーテンプレートまたはよりファットなテンプレートーテンプレートに展開するために用いられるバギー・ルール。

【請求項 6】 上述の抽出ルールが常に一組のシンボル、すなわち $\{s_1, s_2, \dots, s_n\}$ に関連付けられ、このシンボルのそれぞれはテンプレートの 1 つまたは複数のノードに割り当てられ、これらの関連付けられたシンボルには 1 つまたは複数の値が割り当てられてなる抽出ルール

【請求項 7】 入力文を抽出ルールが埋め込まれたテンプレートーテンプレートと照合するための HCS マッチング・アルゴリズム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、抽出ルール・ベースのテンプレートーテンプレート構造およびバギ

ー・ルールの拡張性に富む機能を利用した新しい対話式学習システムに関する。

【0002】

本発明は、言語指向インテリジェント学習システムに使用される時間の掛かるオーサリング・タスクの自動化と簡素化が切望されていることが動機となっている。何故これが切望されているかと言えば、想定される模範回答数は合理的に制限するとしても、理想的な学習システムを開発しようとする場合、学習者が実際に犯す可能性のあるエラーは理論的には無限個に近い数にのぼる場合がしばしばあるからである。本発明者らが判断する限り、少なくとも予見できる将来においては、最先端の自然言語処理技術をもってしても、完全な自由形式でしかもエラーの多い文章の自動訂正を即座に提供出来るレベルには達していない。多くの有能な人間教師のようにシステムがこのような対処ができるには、いわゆる常識と呼ばれる世界の知識ベースをこのシステムに導入することができる場合にのみ可能であろう。

【0003】

【従来の技術および発明の解決すべき課題】

本発明の学習システム「Azalea」にはテンプレート・オートマトンの概念が導入されており、「正しい」回答と「誤った」回答からなる多くの「多様な学習者の予想される例」を収集する。言語学習システムで使われる効率的なエラー診断エンジンとしてHCS（最重共通文字列）またはLCS（最長共通文字列）アルゴリズムという代表的なNLP技術が決定的な役割を担っており、テンプレート内に埋め込まれたこれらの例は学習者の回答の診断解析のために使用される。この診断は、テンプレート・データベースの膨大な数の候補パスの中から学習者の入力文に最も近似度の高いパスを選択することによって行われる。適格なモデル翻訳文と不適格な誤った文章から構成されるテンプレート・コーパスを構築するオーサリング・タスクは、多くの時間を費やし、人手を浪費させる。

【0004】

【課題を解決するための手段】

この新しいシステムには、通常であれば時間の掛かるテンプレート生成のオーサリング・タスクを簡素化、すなわち低減することにおいてのみならず（徳田

尚之、陳 亮、笹井 紘幸らによる特開 2002-49617 号等を参照のこと）、システム性能の向上においても有効である。導入されたテンプレート・テンプレート・アーキテクチャがシステムの簡素化と性能の向上をもたらす第 1 の理由は、このアーキテクチャが、単一のテンプレート・テンプレートの遷移ノードのいくつかに割り当てられた抽出ルールを適用することによって、単一のテンプレート・テンプレートに多くの異なるテンプレートを統合したり、また逆にいうところから多くの異なるテンプレートを同様に抽出することが可能になることであり、第 2 の理由は、導入されたバギー・ルールが学習者の誤った回答を自動的に識別・分類し、したがってそこからバグを生成する機能を有することである。NLP（自然言語処理）技術がこのシステム開発で主要な役割を担う理由は、学習者の自由形式回答の構文構造を調べるためにはパーサが使用されるており、その意味論的構造は、学習者からの回答を準備されたテンプレート・データベースの意味論的に等価のパスとの文字列照合により調べることから明らかであろう。

新しい抽出ルール・ベースの、またバギー・ルール・ベースのテンプレート・テンプレート構造は、文字入力および会話方式による学習システム、音声利用技術によるコール・センターまたは音声ポータル・システム、あるいはシステムと人間との間でより自然な人間とコンピュータの対話を実施するより強化されたヒューマン・コンピュータ・インターフェースに焦点を合わせたシステムならばどのシステムをも含めて多くのアプリケーションで重要な役割を担うことが期待される。

【0005】

この新しい発明では以下のような機能が与えられる。

1. ある制約条件の下で選択したテンプレート・テンプレート遷移図のノードに抽出ルールを埋め込むと、単一テンプレート・テンプレートは様々な異なるタイプの既存のテンプレートを表現することができる。
2. 導入されたバギー・ルールの拡張力に富む機能を利用することによって、多くの誤り表現および／または多くの誤った構文構造は、バグ分類学を記述するその自動機能によって、新しいテンプレート・テンプレートを「よりファットな」テンプレート・テンプレート形式に展開できるように整合性を保って自動的に記

述することができる。語学教師は、テンプレートーテンプレートを制作する際に誤った翻訳文の分類に費やす必要はなく、したがってエラー分類学の負担を低減し、学習者らの特定のエラーの特徴を評価しクラスタ化する。

3. HCS マッチング・アルゴリズムは、入力文をより簡素なテンプレートーテンプレートに直接照合し、したがって実際にテンプレートーテンプレートを展開せずにすべての抽出されたテンプレートのすべての予想されるパスから最適なパスを探索する際の照合処理の空間および時間計算量を低減するように開発することができる。

【0006】

【発明の具体的な態様】

テンプレートーテンプレート構造

まず「テンプレートーテンプレート」という用語を以下の様に定義する。テンプレートーテンプレートは、ノードのいくつかが抽出ルールに関連付けられたシンボルでマーク付けされており、そのテンプレートーテンプレートが多くのテンプレート、または一組の非連結型テンプレートが1つのテンプレートとみなされる場合はいわゆる大きいテンプレートに展開することができる場合の特別なテンプレートと定義される。このような一組の切断されたテンプレートによって、単一のL1文章の様々な予想される翻訳文は、一群の翻訳されたL2文章から構成される大きな単一のテンプレートーテンプレートを形成することが可能になる。展開テンプレートであるので、テンプレートーテンプレート・スキームによってテンプレートーテンプレートはいわば1つまたは複数のテンプレートを抽出することが可能になる。

【0007】

通常、抽出ルールは常に一組のシンボル、例えば $\{s_1, s_2, \dots, s_n\}$ に関連付けられており、そのシンボルのそれぞれにはテンプレートの1つまたは複数のノードが割り当てられている。それらの関連付けられたシンボルには1つまたは複数の値が割り当てられており、その機能は、テンプレートーテンプレートから抽出された1つ以上のテンプレートに現れるノードのスタイルを表現することである。本発明ではこれらのシンボルを「ラベル・シンボル」と呼ぶ。単一

ールに關係付けられたシンボルは「關係シンボル」と呼ばれる。關係シンボルはある種の制約を有するべきである。一般的な制約としては、所与の $s_i = 1$ に際して、 s_k はしばしば 2 に制約されるか、あるいは 1 以外のいくつかの正の整数に制約される必要がある。 s_i の値が一組の他のシンボルに割り当てられた値に依存する場合、 s_i の値の選択は他のシンボルの必須選択肢と呼ばれる。

【0008】

語学教師が容易に理解できるように抽出ルールのいくつかの例を以下に示す。

タイプAルール AP (出現) - NAP (非出現) ルール

ノードによって AP_i でマーク付けされているノードと、 NAP_i でマーク付けされているノードがあると想定する (i は任意の整数で別のタイプのAルールを示す)。タイプAルールの AP - NAP ルールは、展開された時に新たに展開されたテンプレートは AP_i でマーク付けされたノードかまたは NAP_i でマーク付けされたノードを含むことができるが、これらの両方を同時に含むことはできないという条件を課す。本発明においては、 AP_i でマーク付けされたノードがテンプレートに現れないという場合を示すために $AP_i = 0$ を使用する。この時、 NAP_i は、 NAP_i でマーク付けされたノードがそのテンプレートに現れるであろうことを意味する 1 である必要がある。したがって本発明では $NAP_i = 1$ は $AP_i = 0$ の必須選択肢であると言えることができる。同様の理由から、 $NAP_i = 0$ の場合、 $NAP_i = 0$ が $AP_i = 1$ の必須選択肢となるように、 AP_i は 1 の値を有する必要がある。

【0009】

タイプBルール PPR (人称代名詞) - PPRP (人称代名詞所有格) ルール

タイプAルールと同様に、テンプレート-テンプレート・ルールは、一組のテンプレートに現れる PPR_i でマーク付けされたノードと $PPRP_i$ でマーク付けされた他のノード (i は任意の整数) は、それぞれに代名詞の自然言語文法が必要とする人称代名詞のフォームと代名詞の人称代名詞所有格のフォームでなければならないという条件を課す。 $PPRP_i$ (または PPR_i) の場合、 PPR_i (または $PPRP_i$) に要求される値は代名詞の自然言語文法によって定義される必要がある。

【0010】

タイプCルール AN (任意数) ルール

タイプCルールは、 AN_i でマーク付けされたノードには任意のポジティブな実数を割り当てることができるという条件を課す。「I have 5 books on Zen」が真ならば、5以外の数はどの数も誤りなので、このルール AN_i はエラー・ノード5に割り当てることができる。

【0011】

タイプDルール CHO (択一) ルール

タイプDルールは、 CHO_{i_1} , CHO_{i_2} , ..., CHO_{i_k} でマーク付けされたテンプレート-テンプレートのすべてのノードの中で、テンプレート-テンプレートから抽出されたどのテンプレートにもただ一組だけのノードしか現れることができないという条件を課す。ここで異なる i は異なるタイプDルールを表している。したがって、 $CHO_{i_j} = 0$ は CHO_{i_j} でマーク付けされたノードは現れないことを意味し、 $CHO_{i_j} = 1$ はその指定されたノードが現在現れていることを意味している。明らかに、1が1つの CHO_{i_k} に割り当てられている場合、0は他のすべての CHO_{i_j} に割り当てられるべきである。

【0012】

テンプレート-テンプレートを拡張するためのバギー・ルール

ここでいうバギー・ルールは、正確な構文規則からの予想される逸脱によって特徴付けられる一般的な構文上の誤り表現のプロダクション・ルールと定義される。具体的に、バギー・ルールの以下のフォームを想定する。

$$H_1 H_2 \dots H_N \rightarrow R_1 R_2 \dots R_M$$

ここで $H_1 H_2 \dots H_N$ は任意のテンプレート-テンプレートの構文上正確なパスを追跡する一組のノードまたは正確な表現の基本的な構成要素またはセグメントを表す一組の文法的な品詞タグを表す。 $R_1 R_2 \dots R_M$ は、正確なフォームが $H_1 H_2 \dots H_N$ である典型的誤り表現を表すノードのセットである。エラーがテンプレート-テンプレートの正確なパスからのずれによって識別されることが即座に理解される。

【0013】

一例を示す。

EX VBP→EX VBZ

(ここでEXは「there is」(～がある)というような存在を表すものであり、PRPは人称代名詞であり、VBPは1人称および2人称の現在形の動詞であり、VBZは3人称単数現在形の動詞である)。この例は、構文上正確な表現である「there are 5 books」が、主語と動詞の属性一致を誤って理解している生徒によって誤用されており、この例では誤り表現である「there is 5 books」が生じたことを意味している。

【0014】

【発明の具体例の説明】

以下本発明を図面に示す具体例によって説明する。

図1はオリジナルなテンプレートーテンプレートの構成を示す図であり、図2は上記の展開ルールにより展開されたテンプレートーテンプレートを示す図、図3はテンプレート1を示す図であり、図4はテンプレート2を示す図である。

【0015】

テンプレートーテンプレート、抽出ルールによるテンプレート拡張、およびバギー・ルールの例

本発明野実施例においては、まず「Japan is dotted with beautiful gardens nationwide.」という意味の日本語の文章の英語への翻訳文に関するテンプレートーテンプレートを構築する。図1等示される番号は、文中における各単語の相対的重要性を強調する各単語の重みを示している。テンプレートの単語のデフォルトの重みは1にセットされており、これらはこの分野の専門家によって判断されるその単語の重要性に関連付けて割り当てられる必要がある。徳田 尚之、陳 亮、笹井 絃幸による特開2002-49617号を詳細な説明のために参照する。「[]」の間のシンボルは品詞タグである。図中左端に示すノードは開始ノードである。

【0016】

前記バギー・ルールを単純に適用することによって、それを図2のテンプレートーテンプレートに展開することができる。

これは、前記のバギー・ルールがバグの分類法を生成することができ、それによってそれらの誤り表現をテンプレート-テンプレートに構築することを自動的に可能にするので、語学教師がテンプレート-テンプレートを構築する際に多くの一般的なエラーを分類することの詳細に配慮する必要はないということを示している。

【0017】

次に、前記タイプAルールを適用すると、 AP_1 でマーク付けされたノードがそのテンプレートに現れることを可能にし、したがって図2の NAP_1 でマーク付けされたノードを削除することによって図2のテンプレート-テンプレートから図3のテンプレート1を抽出し、ならびに図2の AP_1 でマーク付けされたノードを削除し、したがって NAP_1 でマーク付けされたノードがそのテンプレートに現れるようにすることによって図4のテンプレート2を抽出することができることが容易に分かる。

語学教師がより簡素なラベル・シンボルに関してテンプレートの1つの大きな組み合わせを完成させてテンプレート-テンプレートを構築することができることが理解される。

【0018】

テンプレート-テンプレートおよび入力文に対する最重共通文字列マッチング・アルゴリズム

前記の詳細な説明から明らかなように、単一テンプレート-テンプレートから多くのテンプレートを抽出することができる。テンプレート-テンプレートが遷移図上のあるノードに関連付けられるべきラベル・シンボル s_1, s_2, \dots, s_n を有すると仮定すると、テンプレート-テンプレートから抽出された異なるテンプレートはそれらのシンボルをノードに割り当てることによって獲得することができる。本発明ではテンプレートから抽出された各テンプレートを n タプル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ で示すことができる。ここで、 p_i はシンボル s_i に対する適切な割り当てである。上記のセクションで既に議論した通り、 p_i は使用される抽出ルールに従って数字または単語のどちらかであってよい。

【0019】

2つの文章の最重共通文字列は a_1 の次に a_2 、その次に...、その次に a_m という順番で両方の文章に現れる順番付けされた単語の最重共通文字列 a_1, a_2, \dots, a_m であると定義される。共通文字列の定義は、A. V. Aho および J. D. Ullman 著 (Computer Science Press, 1992, pp. 321-327) の教科書『Foundations of Computer Science [情報科学の基礎]』に記載されている。

テンプレート内の各単語または各語句には重みが割り当てられているので、テンプレート内のパスと入力文の最重共通文字列は、その合計の重みが最大であるすべての予想可能な共通文字列内の最重共通文字列と定義される。

【0020】

次に、テンプレートのすべての予想可能な有効パスから入力文の単語および／または語句の最重共通文字列を検索する。

テンプレートおよび入力文の最重共通文字列は、それぞれがテンプレートと入力文1つのパスから獲得される最重共通文字列中で最も重い合計重量を有する単語の最重共通文字列と定義される。

言語翻訳学習システムに関するアプリケーションでテンプレートーテンプレートが獲得されると、次のステップは、入力文をすべての予想可能なテンプレートのそれぞれと照合し、次いで最も近いパスを選択することである。文章に対するテンプレートの DP (ダイナミックプログラミング) ベースの照合手順の詳細な説明は、徳田 尚之、陳 亮、笹井 紘幸らによる特開 2002-49617 号等に記載されている。

【0021】

本発明の方法では、テンプレートーテンプレートからすべてのテンプレートを物理的に抽出せずに、抽出ルールによって（しかしバギー・ルールにはよらずに）直接的に抽出することができるテンプレートのすべての有効なパスの中から最も近いパスを探索する。そのような照合を行う前にテンプレートーテンプレートにはいかなるバギー・ルールをも含めないように、まずバギー・ルールが埋め込まれたテンプレートーテンプレートを展開する必要がある。これはすでに述べた

図2のステップに従って実行することができる。

【0022】

アルゴリズムで必要とされる第1のステップは、適用可能な空ノードごとに□というラベルの弧に重み0を加算してテンプレートの各ノードをグラフの1つまたは複数の弧として単純に表現することによって、テンプレート→テンプレートを非循環的な重み付けされた有限有向グラフ（有向グラフ）の2重数値に変換することである。有向グラフはテンプレート→テンプレートから変換されるので、その関数とそのシンボルに割り当てられた値に決定的に依存するラベル・シンボルに関連付けられた多くの弧を含む。したがってそのような1つの有向グラフがあると仮定すると、異なる一組のラベル・シンボルが弧に割り当てられている場合は完全に異なるテンプレートを抽出することができる。すなわち、そのような有向グラフがあると仮定すると、テンプレート→テンプレートから抽出することができるテンプレートにそれぞれが対応する多くの有向グラフを獲得することができる。テンプレート→テンプレートから抽出された有向グラフを、以後、テンプレート有向グラフと呼ぶ。

【0023】

本発明者らは、次にすべての有向グラフのパスと入力文の共通文字列から最重共通文字列を探索する手順を以下で定義する。

有向グラフの任意の特定のノードNで終了するパスと入力文の最重共通文字列は、有向グラフNで終了する1つのパスと入力文からそれぞれが獲得されるすべての最重共通文字列中最も重い合計重量を有する単語のシーケンスと定義される。

さらに、テンプレート有向グラフから抽出され、しかしノード N_i で終了するすべての有向グラフのパスを $N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ で表現する。ここでシンボル s_i は値 p_i ($i=1, 2, \dots, n$)に割り当てられる。 n タプル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ をノード N_i のラベルと呼ぶ。ここで s_1 を p_1 に、 s_2 を p_2 に、 \dots 、 s_n を p_n にセットするときルールに矛盾がないものと仮定すべきである。このようなラベル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ を矛盾なしラベルと呼ぶ。

$N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ でラベルされたノードと入力文の最重共通文字列は、矛盾なしラベル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ でマーク付けされたノードを有する有向グラフ・テンプレートから抽出された1つの有向グラフの最重共通文字列としてそれぞれが獲得されるすべての最重共通文字列中で最も重い合計の重みを有する単語の最重共通文字列と定義される。

【0024】

ノードによっては、1つの有向グラフのAP2でラベルされたノードとNAP2でラベルされたノードのように有向グラフ・テンプレートから抽出された1つの有向グラフで同時に現れない場合があることに留意されたい。その結果、 N_i ($\dots, AP2, 1, \dots, NAP2, 1 \dots$) のようにルール違反のラベルは、有向グラフ・テンプレートのノードと入力文の共通文字列のどの計算計画にも入ることを許可されるべきではない。以下のアルゴリズムはテンプレート→テンプレートと入力文の最重共通文字列を計算するための手順を記述する。以下の計算では、ラベル・シンボルの非常に特別の値として「 λ 」が使用され、それによってその値は計算の特定の段階に達するまで未指定のままである。

【0025】

1. テンプレート→テンプレートを、その有向エッジ（遷移）がテンプレートの対応する単語によってラベルされるテンプレート有向グラフに変換する。
2. 有向グラフのすべてのノードを、ノード N_i および N_j の対ごとに $j > i$ の場合に N_j から N_i への遷移がないように N_1, N_2, \dots, N_t のノードにトポロジカリーに分類する。
3. 空のノード N_0 を有向グラフに追加し、 N_0 からの弧をテンプレート→テンプレートのすべての開始ノードに加える。
4. $CM(N_0, M_0) = 0$ にセットする。
5. $i = 0$ から t の場合、以下のステップを実行する。

【0026】

6. 1つのシンボルに関連付けられた N_i に対して少なくとも1つの弧があり、ノード N_i 以降に s の関係ラベルがまったく存在しない場合、すべての i に関し

て $j = 0$ から m の場合は以下を実行する。

s 関係ラベルが $\{s_1, s_2, \dots, s_n\}$ に現れておらず少なくとも1つの $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s_{x_1}, p_{x_1}, s_{x_2}, p_{x_2}, \dots, s_{x_h}, p_{x_h}\}$ が存在する場合、任意のラベル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ に関してすべての $CM(N_i \{ \dots \}, M_j)$ をチェックする。ここで $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s_{x_1}, p_{x_1}, s_{x_2}, p_{x_2}, \dots, s_{x_h}, p_{x_h}\}, M_j)$ が定義されるように、 $s_{x_1}, s_{x_2}, \dots, s_{x_h}$ は s 関係ラベルである。

$CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ を最大の $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s_{x_1}, p_{x_1}, s_{x_2}, p_{x_2}, \dots, s_{x_h}, p_{x_h}\}, M_j)$ と定義し、すでに定義されているすべての $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s_{x_1}, p_{x_1}, s_{x_2}, p_{x_2}, \dots, s_{x_h}, p_{x_h}\}, M_j)$ を定義解除する。

【0027】

7. $j = 0$ から m の場合、以下のステップを実行する。

8. N_i からの弧があるノード N_k のそれぞれに対しては以下を実行する。

(1) 弧 $N_i N_k$ がラベルを有しない場合、すでに定義されている $CM(N_i \{ \dots \}, M_j)$ 、 $CM(N_i \{ \dots \}, M_{j+1})$ 、 $CM(N_k \{ \dots \}, M_j)$ 、 $CM(N_k \{ \dots \}, M_{j+1})$ をすべてチェックし、 $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ 、 $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$ 、 $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ 、 $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$ の一つが既に定義されている場合は $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$ を以下のデータの最大と定義する。

・ 既に定義されている場合は $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ である。

・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ が既に定義されており、弧 $N_i N_k$ が W_k と一致する場合は $M(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j) + W(N_i N_k)$ である。

・ 既に定義されている場合は $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$ である。

・ 既に定義されている場合は $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ である。

既に定義されている場合は $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$ である。

【0028】

(2) 弧 $N_i N_k$ がシンボル s に関連付けられている場合、既に定義されている $CM(N_i (...), M_j)$ 、 $CM(N_i, \{...\}, M_{j+1})$ 、 $CM(N_k (...), M_j)$ 、 $CM(N_k \{...\}, M_{j+1})$ をチェックする。

(i) ノード・ラベル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}$ が矛盾なしラベルであり、以下の少なくとも1つが既に定義されている場合：

・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ 、 $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$

・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ 、 $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_{j+1})$

・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_j)$ 、 $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_{j+1})$

・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_j)$ 、 $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_{j+1})$

・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, \lambda\}, M_{j+1})$ を上記で定義されたデータの最大のデータと定義する。

【0029】

(ii) ラベル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ があると仮定して、ラベル $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}$ が矛盾なしラベルであり、以下の少なくとも1つが真である場合：

・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ がすでに

定義されており、 s から p までのセッティングが $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ の必須選択肢であるか、 p が s に割り当てられた後で $N_i N_k$ が M_{j+1} と一致する場合

- ・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\} M_{j+1})$ がすでに定義されており、 s から p までのセッティングが $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ の必須選択肢の場合

- ・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ がすでに定義されており、 s から p までのセッティングが $\{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}$ の必須選択肢である場合

- ・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_{j+1})$ が既に定義されている場合

- ・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_j)$ が既に定義されている場合

- ・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, , s, p\}, M_{j+1})$ が既に定義されている場合

- ・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, , s, p\}, M_j)$ が既に定義されている場合

- ・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, , s, p\}, M_{j+1})$ が既に定義されている場合

- ・ $CM(N_k \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_{j+1})$ を上記で定義されたデータと以下のデータの最大データと定義する。

【0030】

- ・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j)$ が定義されており、 p が s にセットされた後で $N_i N_k$ が M_{j+1} に一致する場合、 $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n\}, M_j) + W(M_{j+1})$ である。

- ・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_j)$ が定義されており、 p が s に割り当てられた後で $N_i N_k$ が M_{j+1} と一致する場合、

・ $CM(N_i \{s_1, p_1, s_2, p_2, \dots, s_n, p_n, s, p\}, M_j) + W(M_{j+1})$ である。

また、 s として割り当てられた p に従い、これらの s_i を λ から値の必須選択肢まで変更する。

【0031】

N_x を最終的な頂点であると既に定義しているすべての $CM(N_x, M_m)$ の中で最大の $CM(N_x, M_m)$ はテンプレートーテンプレートおよびパスの最重共通文字列の重みになる。

上記のアルゴリズムでは、いくつかの候補からいつ $CM(N, (..), M)$ を選択したとしても選択したその1つに対して一種のバック・リンクをセットすることに留意されたい。テンプレートーテンプレートとパスの最長共通文字列の重みを探索したように、このバック・リンクを追跡することによって、最重共通文字列を有する抽出されたテンプレートのパスを入力文と共に即座に獲得することができることに留意されたい。

【0032】

【発明の効果】

自然言語学習システムでの必要性が本発明の動機となっており、したがって本願明細書中では自然言語翻訳の技術分野に関して記載したが、本発明の用途は自然言語学習システムに限定されるものではない。すなわち、本発明は音声利用技術、プログラミング言語学習システム、または例えば人間とコンピュータの対話を可能にするより自然な拡張型インターフェースを必要とするシステムならばどのシステムのためにでも使用することができる。

【図面の簡単な説明】

【図1】

本発明のテンプレートーテンプレートの構造を示す図である。

【図2】

展開ルールにより展開されたテンプレートーテンプレート構造を示す図である。

。

【図3】

「Japan is dotted with beautiful parks nationwide」という意味の文章の例で展開されたテンプレート 1 を示す図である。

【図 4】

「Japan is dotted with beautiful parks nationwide」という意味の文章の例で展開されたテンプレートを示す図である。

【符号の説明】

エラー・メッセージ:

AS: 名詞の量に仮定が行われている、

AT: 冠詞は必要ない、

CM: コンマが必要である、

CT: 短縮形が不正確である、

MN: 意味が不正確である、

NP: 名詞は複数形でなければならない、

VS: 主語が単数形なので動詞は単数形でなければならない、

PR: 前置詞が不正確である、

PP: 句は複数形である必要がある、

一般的な品詞タグ:

DT: 限定詞

EX: 存在を表す

IN: 前置詞／従属接続詞

JJ: 形容詞

NN: 名詞、単数または集合

NNS: 名詞、複数

NNP: 固有名詞、単数

RB: 副詞

VBN: 動詞、過去分詞

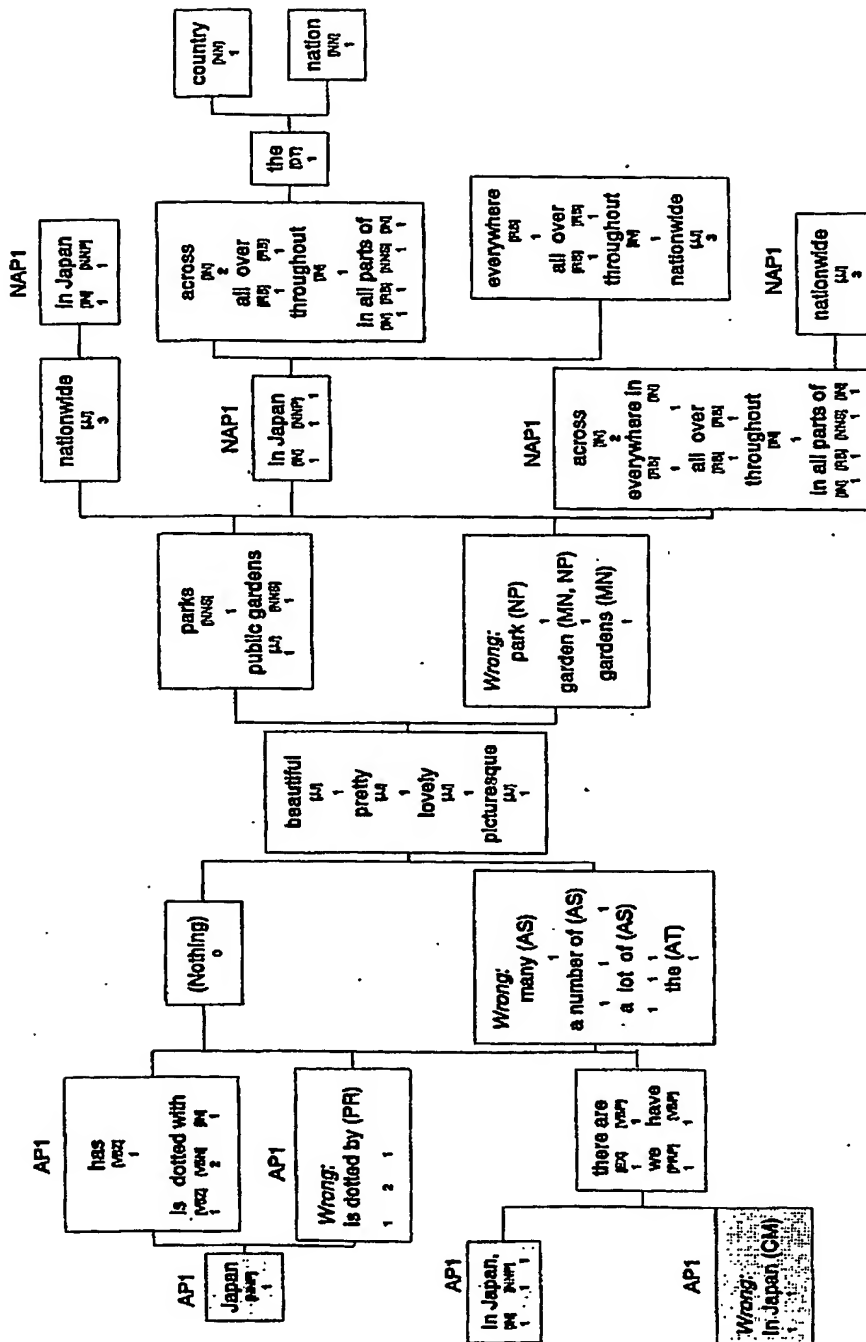
VBP: 動詞、非 3 人称単数、現在形

VBZ: 動詞、3 人称単数、現在形

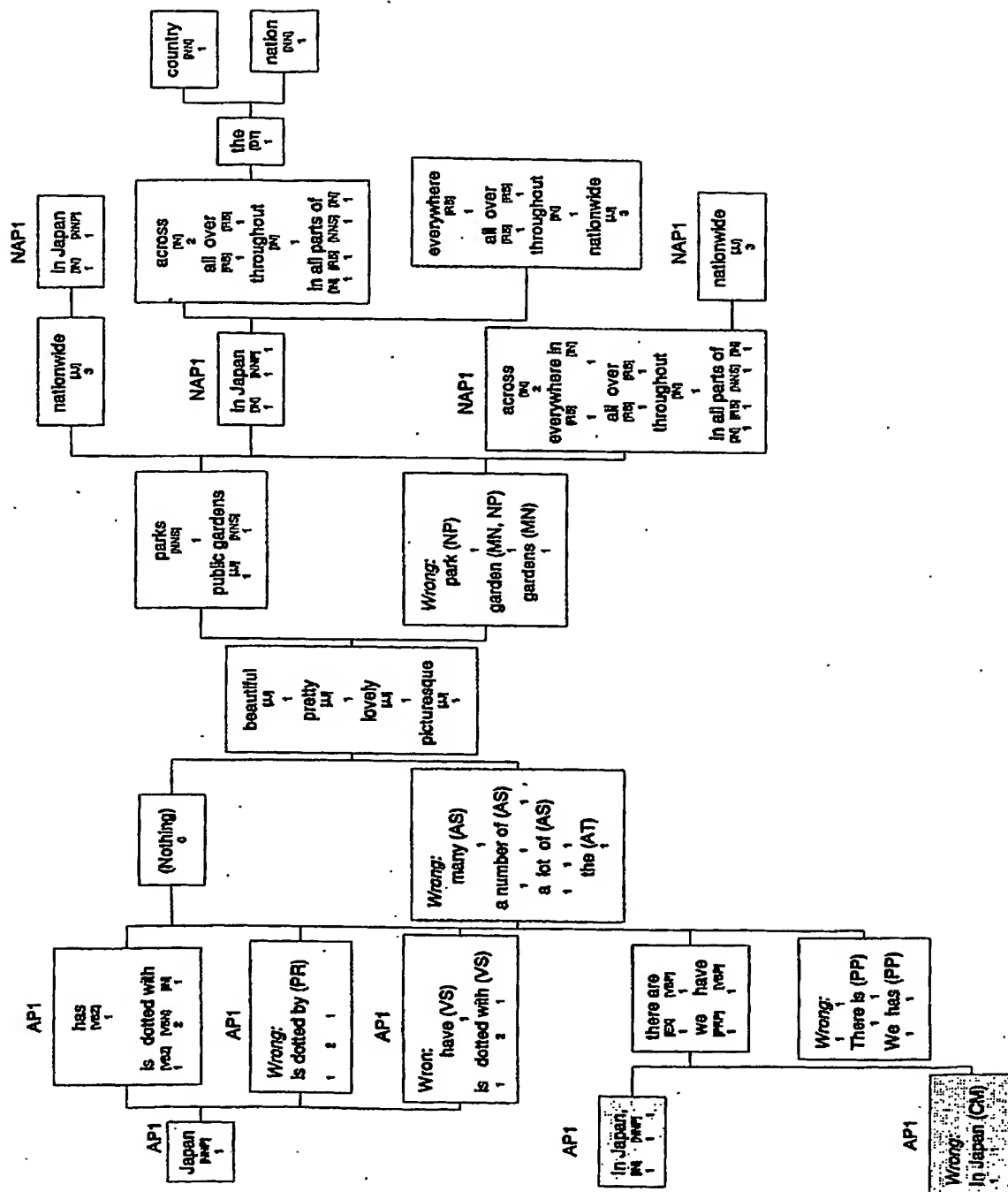
【書類名】

凶面

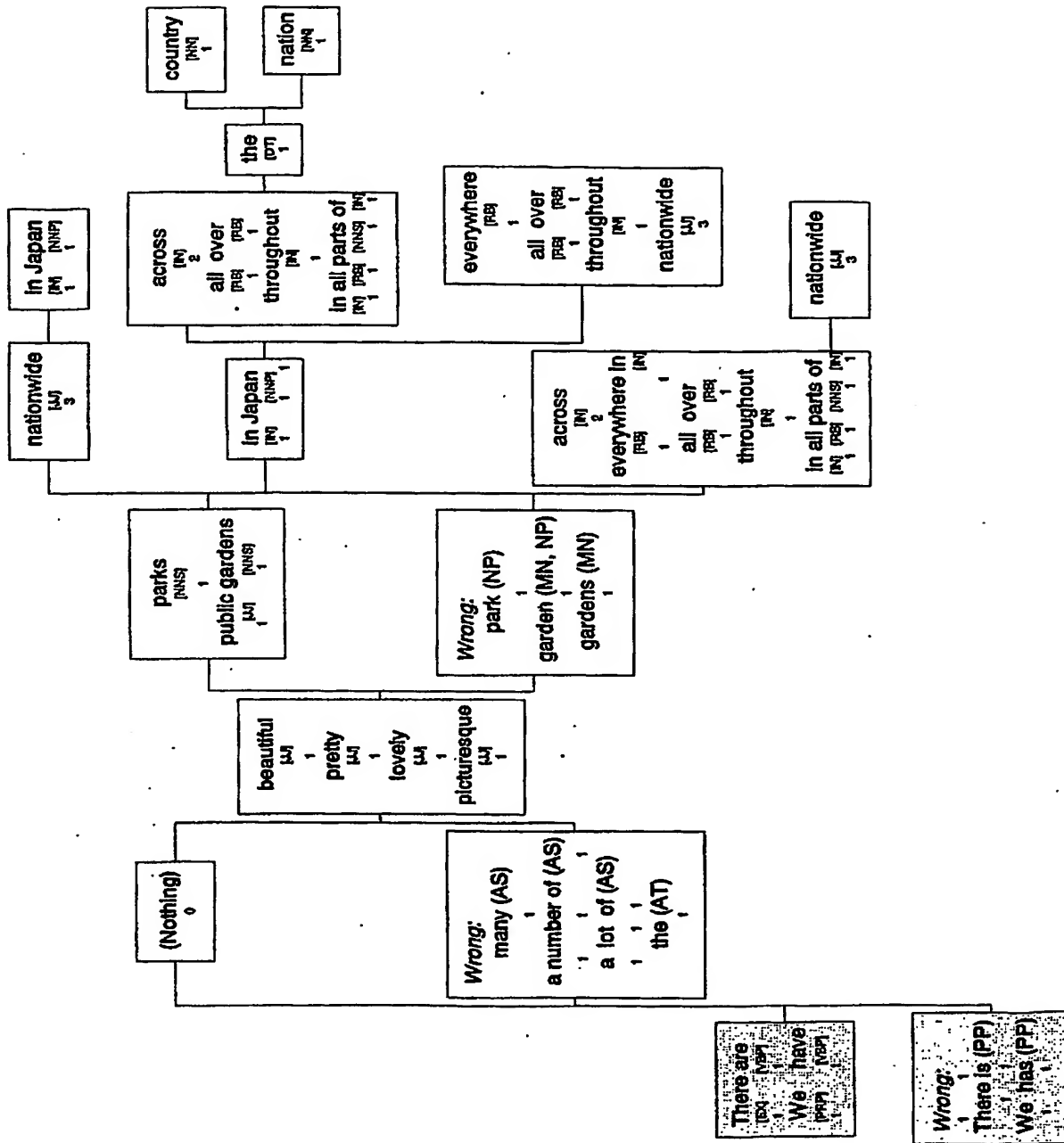
【図 1】



【図 2】



【図 4】



【書類名】 要約書

【要約】

【解決すべき課題】

言語指向インテリジェント学習システムに使用される時間の掛かるオーサリング・タスクの自動化と簡素化を実現する。

【課題を解決するための手段】

本発明の学習システム「Azalea」にはテンプレート・オートマトンの概念が導入されており、「正しい」回答と「誤った」回答からなる多くの「多様な学習者の予想される例」を収集する。言語学習システムで使われる効率的なエラー診断エンジンとしてHCS（最重共通文字列）またはLCS（最長共通文字列）アルゴリズムという代表的なNLP技術が決定的な役割を担っており、テンプレート内に埋め込まれたこれらの例は学習者の回答の診断解析のために使用される。この診断は、テンプレート・データベースの膨大な数の候補パスの中から学習者の入力文に最も近似度の高いパスを選択することによって行われる。

【発明の効果】

新しい抽出ルール・ベースの、またバギー・ルール・ベースのテンプレート・テンプレート構造は、文字入力および会話方式による学習システム、音声利用技術によるコール・センターまたは音声ポータル・システム、あるいはヒューマン・コンピュータ・インターフェースに焦点を合わせたシステムをも含めて多くのアプリケーションで重要な役割を担うことが期待される

認定・付加情報

特許出願の番号	特願 2003-120733
受付番号	20300530358
書類名	特許願
担当官	田口 春良 1617
作成日	平成15年 6月18日

<認定情報・付加情報>

【特許出願人】

【識別番号】	596091749
【住所又は居所】	東京都杉並区西荻南 4-3-2
【氏名又は名称】	川井 忠彦

【代理人】

申請人

【識別番号】	100082153
【住所又は居所】	東京都渋谷区代々木 2 丁目 23 番 1-1144 号 小原・駒津特許事務所

【氏名又は名称】	小原 二郎
----------	-------

【選任した代理人】

【識別番号】	100080078
【住所又は居所】	東京都渋谷区代々木 2-23-1-1144 小 原・駒津特許事務所

【氏名又は名称】	駒津 敏洋
----------	-------

次頁無

【書類名】 手続補正書（方式）

【提出日】 平成15年 6月10日

【整理番号】 P0150322

【あて先】 特許庁長官殿

【事件の表示】

【出願番号】 特願2003-120733

【補正をする者】

【識別番号】 596091794

【氏名又は名称】 株式会社サン・フレア

【代理人】

【識別番号】 100082153

【弁理士】

【氏名又は名称】 小原 二郎

【発送番号】 051737

【手続補正 1】

【補正対象書類名】 特許願

【補正対象項目名】 特許出願人

【補正方法】 変更

【補正の内容】

【特許出願人】

【識別番号】 596091794

【住所又は居所】 東京都新宿区四谷4丁目7番地

【氏名又は名称】 株式会社サン・フレア

【プルーフの要否】 要

認定・付加情報

特許出願の番号	特願 2003-120733
受付番号	50300969713
書類名	手続補正書 (方式)
担当官	田口 春良 1617
作成日	平成15年 6月18日

<認定情報・付加情報>

【補正をする者】

【識別番号】

596091794

【住所又は居所】

東京都新宿区四谷4丁目7番地

【氏名又は名称】

株式会社サン・フレア

【代理人】

申請人

【識別番号】

100082153

【住所又は居所】

東京都渋谷区代々木2丁目23番1-1144号

小原・駒津特許事務所

【氏名又は名称】

小原 二郎

次頁無

特願 2 0 0 3 - 1 2 0 7 3 3

出 願 人 履 歴 情 報

識別番号

[5 9 6 0 9 1 7 4 9]

1. 変更年月日

1 9 9 6 年 7 月 1 2 日

[変更理由]

住所変更

住 所

東京都杉並区西荻南 4 - 3 - 2

氏 名

川井 忠彦

特願 2 0 0 3 - 1 2 0 7 3 3

ページ : 2/E

出 願 人 履 歴 情 報

識別番号

[5 9 6 0 9 1 7 9 4]

1. 変更年月日
[変更理由]

住 所
氏 名

1 9 9 9 年 1 月 1 9 日

住所変更

東京都新宿区四谷 4 丁目 7 番地

株式会社サン・フレア

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.